

## A big data kihívás a bölcsészettudományokban: néhány digitális bölcsészeti kutatási eszköz bemutatása

**Péter Róbert**

egyetemi docens

Szegedi Tudományegyetem Bölcsészet- és Társadalomtudományi Kar, Angol-  
Amerikai Intézet, Angol Tanszék

[robert.peter@ieas-szeged.hu](mailto:robert.peter@ieas-szeged.hu)

### Absztrakt

Az előadás célja, hogy bemutasson néhány, digitális bölcsészeti kutatások során használt, jól ismert eszközt (pl. Google N-Gram Viewer, Bookworm, Voyant, Juxta Commons), valamint a Szegedi Tudományegyetemen az elmúlt évben kifejlesztett – és fejlesztés alatt lévő – két webszolgáltatást (TANIT és AVOBMAT). A TANIT (Text ANalysIs Tools) rendszer célja, hogy magyar nyelvű szövegek számítógépes nyelvészeti feldolgozásával dokumentumok összehasonlító elemzéséhez szükséges statisztikákat kigyűjtsön. Ez a webszolgáltatás létező nyelvtechnológiai elemzőlánc kimenetére épülő aggregált statisztikákat számít ki, témamodelleket épít, és ezeket olyan formátumban adja át a digitális bölcsész felhasználónak, aki utána programozói ismeretek nélkül is fel tudja ezt használni kutatásaiban. Az AVOBMAT (Analysis and Visualization of Bibliographic Metadata and Texts) segítségével nagy mennyiségű metaadatot és szöveget tudunk elemezni és vizualizálni. Az AVOBMAT-ba saját fájlokat és könyvtári repozitóriumokat tölthetünk fel, többek között Zoteróból exportált csv és rdf, valamint EP3 xml formátumokban. A feltöltés után tudjuk az adathalmazt szűkíteni fazettás, összetett és CCL kereséssel. A metaadatokat számtalan módon tudjuk interaktív módon vizualizálni, amelynek segítségével új, eddig ismeretlen összefüggéseket és trendeket fedezhetünk fel digitális bölcsészeti elemzések során. Az AVOBMAT az egyszerű vizualizációk segítségével tudja például modellezni szerzők, kiadók, kulcsszavak kapcsolatát és időbeni eloszlását. A mesterséges intelligenciás módszereket és technológiákat használó összetett vizualizációs funkciók között szerepel a szerző–kiadó–(eladó) hálózatok interaktív vizualizációja, valamint női–férfi szerzők automatikus azonosítása és modellezése 55 nyelven. Az AVOBMAT működését a szegedi Acta repozitórium és brit sajtócikkek elemzésével fogom demonstrálni.

**kulcsszavak:** digitális bölcsészet, big data, TANIT, AVOBMAT